SASを用いたC-vineコピュラによる擬似乱数の生成

矢田真城[1]　魚住龍史[2]　浜田知久馬[3]

[1]エイツーヘルスケア株式会社　データサイエンス第2本部　生物統計部
[2]京都大学大学院医学研究科　医学統計生物情報学
[3]東京理科大学　工学部　経営工学科

Generating pseudo-random numbers from C-vine copulas by using SAS

Shinjo Yada[1] , Ryuji Uozumi[2], and Chikuma Hamada[3]

[1]Biostatistics Department, A2 Healthcare Corporation
[2]Kyoto University Graduate School of Medicine
[3]Department of Management Science, Tokyo University of Science

要旨:

C-vineとは，同時確率密度関数を，周辺確率密度関数と適当な条件付き分布関数を引数とするペアコピュラを用いて構成する方法の1つである.

本発表では，C-vineに基づく擬似乱数生成法について紹介する.

キーワード：Vine copula, regular vine, C-vine,擬似乱数

コピュラとは

- コピュラ（Copula）

  1次元周辺分布を接合して多次元同時分布を構成するときの接合関数のこと

  - 様々なタイプのコピュラ
    - 楕円コピュラ，アルキメディアンコピュラ
  - COPULAプロシジャ
    - 擬似乱数の生成，コピュラパラメータの推定
    - SAS/ETS 13.1より正規版に

多次元分布のモデル化

使いやすいコピュラが少ない

- よく知られているアルキメディアンコピュラは，少数のコピュラパラメータで構成される．
- 楕円コピュラでは，分布の裾における変数間の従属関係に，ある種の制約がある．

➡ 2次元のコピュラを用いて，3次元以上の同時分布を構成しようというアプローチ

同時確率密度関数の分解

- 確率変数X_1, X_2, X_3の同時確率密度関数

$$f(x_1, x_2, x_3) = f_1(x_1)\, f_{2|1}(x_2|x_1)\, f_{3|12}(x_3|x_1,x_2)$$

$F_{12}(x_1, x_2) = C_{12}(F_1(x_1), F_2(x_2))$ を満たすコピュラC_{12}に対して

$c_{12}(u,v) = \dfrac{\partial^2}{\partial u \partial v} C_{12}(u,v)$ （pair copula density function）とおくと

$$f_{2|1}(x_2|x_1) = c_{12}(F_1(x_1),F_2(x_2)) \times f_2(x_2)$$

$$f_{3|12}(x_3|x_1,x_2) = c_{23|1}(F_{2|1}(x_2|x_1),F_{3|1}(x_3|x_1)) \times c_{13}(F_1(x_1),F_3(x_3)) \times f_3(x_3)$$

$$\boxed{\begin{aligned} f(x_1, x_2, x_3) = {} & f_1(x_1)\, f_2(x_2)\, f_3(x_3) \\ & \times c_{12}(F_1(x_1), F_2(x_2)) \times c_{13}(F_1(x_1), F_3(x_3)) \\ & \times c_{23|1}(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)) \end{aligned}}$$

同時確率密度関数の分解

- ペアコピュラと周辺分布との積

同時確率密度関数 = 1次元周辺確率密度関数の積

×適当な分布関数を引数とするペアコピュラ

  - ペアコピュラが一意に定まらない
  - 多次元になるほど数多くのペアが存在する

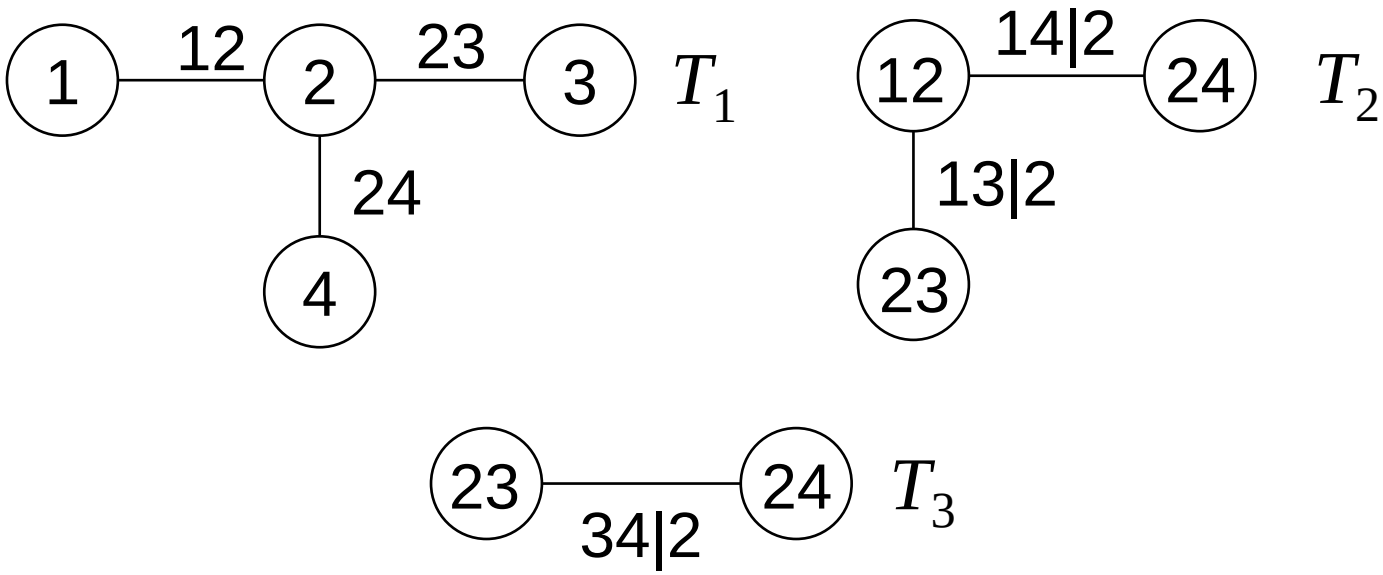
Vine copula（ヴァインコピュラ）

＝vineと呼ばれるグラフィカルモデルに基づくcopula

Vine

- グラフィカルモデルの1つ

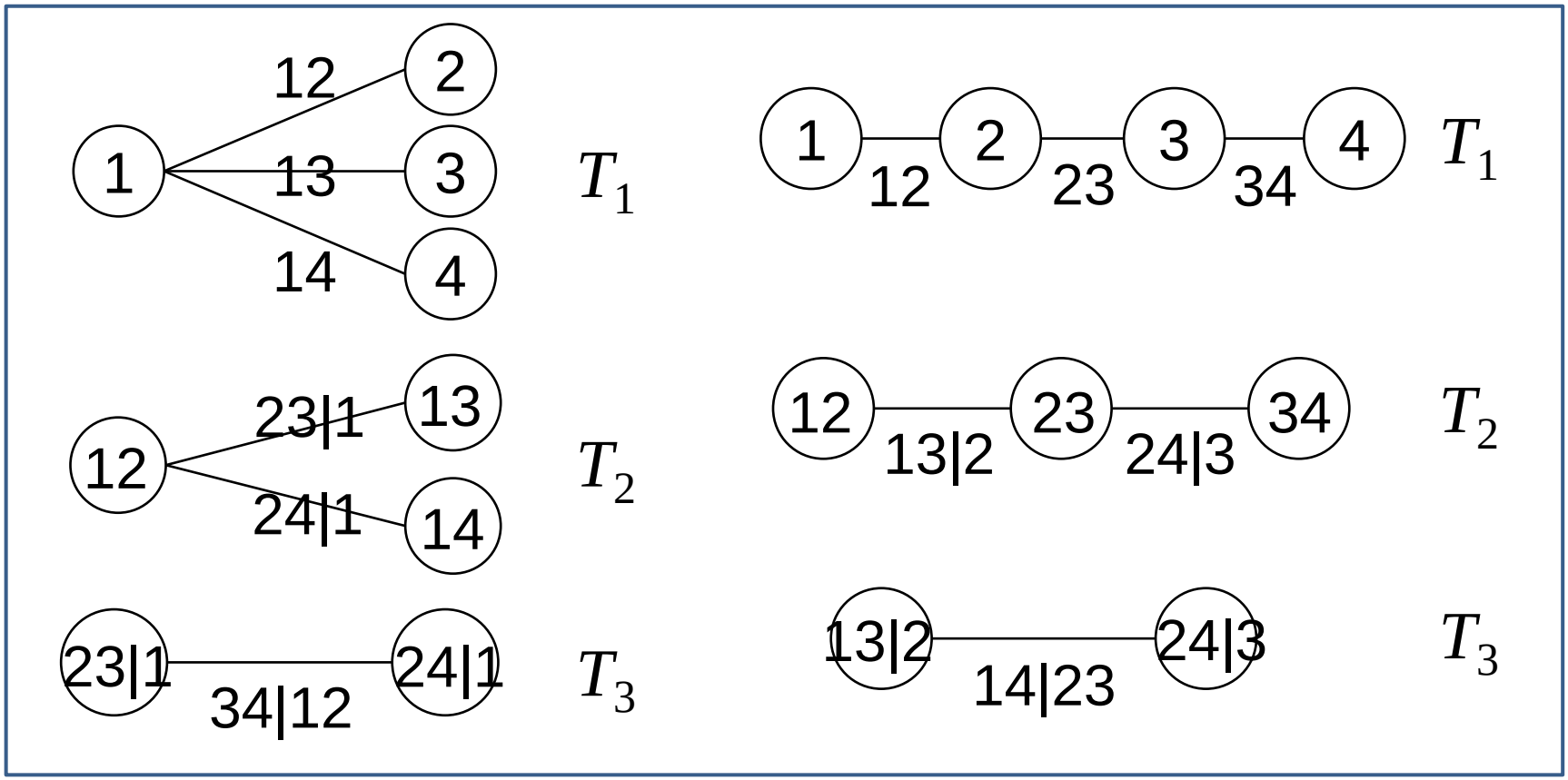
  確率変数間の依存構造を, 頂点(node)と辺(edge)からなる木(tree)により表現したグラフ



- Vineの代表格がRegular vine（正則ヴァイン）

Regular-vine

- Canonical-vine（C-vine）とD-vine



4変数のC-vine（左）とD-vine（右）

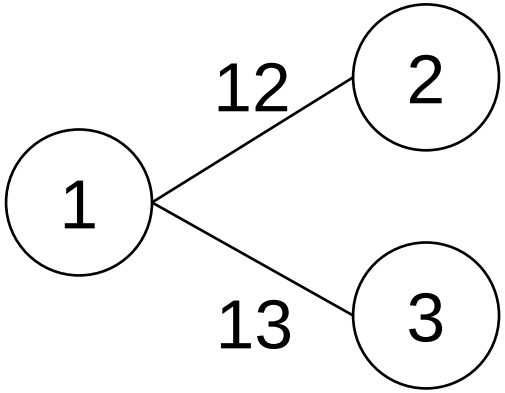
C-vine copula

重要な変数を1つ選び，それ以外の変数との関係をペアコピュラでモデル化する.

- 3変数の場合

  3つの確率変数X_1, X_2, X_3のうち，X_1が重要な変数である場合について考える.

  1) X_1とX_2の関係，X_1とX_3の関係をモデル化する.



T_1

$N_1=\{1,2,3\}$

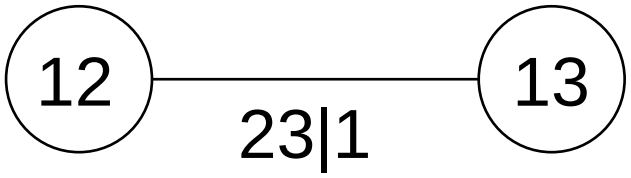
$E_1=\{e_1=\{1,2\},e_2=\{1,3\}\}$

C-vine copula

- 3変数の場合（続き）

2) T_2を構成する頂点のうち，いずれか1つをT_2の根として選択し，根とそれ以外の頂点の関係を，1つ前の木の根で条件付けてモデル化する.

- T_2を構成する頂点$=\{1,2\},\{1,3\}$
- 1つ前の木の根$:1$



T_2

$N_2=\{\{1,2\},\{1,3\}\}$

$E_2=\{\{e_1,e_2\}\}$

C-vine copula

- 3変数の同時確率密度関数

  “3次元同時確率密度関数を,
  適当な分布関数を引数とするペアコピュラと,
  1次元の周辺確率密度関数を用いて構成する”
  というアプローチ

$$f(x_1, x_2, x_3) = f_1(x_1)\, f_2(x_2)\, f_3(x_3)$$

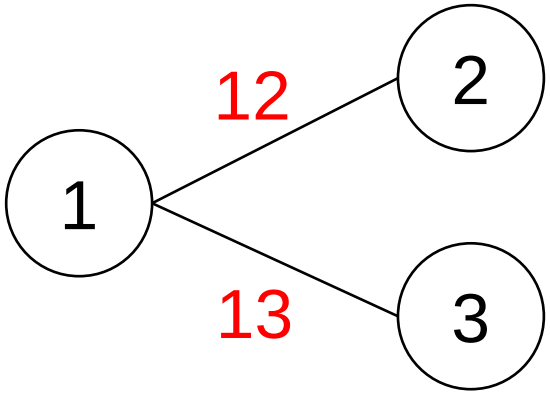
× <u>適当な分布関数を引数とするペアコピュラ</u>

先程のグラフから書きくだすことができる！

C-vine copula

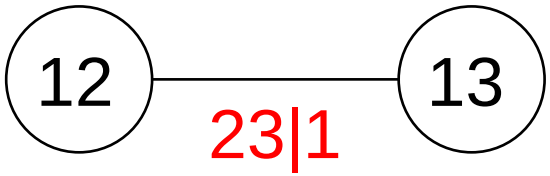
- 3変数の同時確率密度関数

  辺がペアとなるコピュラであり，辺に付与したラベルがペアコピュラ密度関数の下付き文字に対応



T_1 ⟺ $c_{12}(F_1(x_1),F_2(x_2))$

$c_{13}(F_1(x_1),F_3(x_3))$



T_2 ⟺ $c_{23|1}(F_{2|1}(x_2|x_1),F_{3|1}(x_3|x_1))$

C-vine copulaからの乱数生成

- コアとなる手順

1) 互いに独立に，区間[0,1]の一様分布に従う乱数 w_1, w_2, w_3を生成する.

2) 1)で生成した一様乱数を用い

$$u_1 = w_1$$

$$u_2 = F_{2|1}^{-1}(w_2 \mid u_1)$$

$$u_3 = F_{3|1,2}^{-1}(w_3 \mid u_1, u_2)$$

条件付き分布関数からの乱数生成が大きな問題

により，1組の乱数(u_1, u_2, u_3)を得る.

C-vine copulaからの乱数生成

- h関数

  条件付き分布関数とコピュラとの関係を示したもの

$$h(x,v,\boldsymbol{\theta}) := F_{x|v}(x\,|\,v) = \frac{\partial C_{x,v}(x,v)}{\partial v}$$

$F_{x|v}(x|v)$: 条件付き分布関数.

$C_{x,v}(x,v)$: 2次元コピュラ. $\boldsymbol{\theta}$ はコピュラパラメータベクトル.

- h逆関数

  h関数の第1引数に対する逆関数

$u = h(x,v,\boldsymbol{\theta})$ より $x = h^{-1}(u,v,\boldsymbol{\theta})$

C-vine copulaからの乱数生成

- h関数による条件付き分布関数の表記

$$F_{2|1}(x_2 \mid x_1) = \frac{\partial C_{2,1}(F_2(x_2), F_1(x_1))}{\partial F_1(x_1)} = h(x_2, x_1, \boldsymbol{\theta}_{\mathbf{1,1}})$$

$$F_{3|1}(x_3 \mid x_1) = \frac{\partial C_{3,1}(F_3(x_3), F_1(x_1))}{\partial F_1(x_1)} = h(x_3, x_1, \boldsymbol{\theta}_{\mathbf{1,2}})$$

$$\begin{aligned} F_{3|1,2}(x_3 \mid x_1, x_2) &= \frac{\partial C_{2,3|1}(F_{3|1}(x_3 \mid x_1), F_{2|1}(x_2 \mid x_1))}{\partial F_{2|1}(x_2 \mid x_1)} \\ &= h(F_{3|1}(x_3 \mid x_1), F_{2|1}(x_2 \mid x_1), \boldsymbol{\theta}_{\mathbf{2,1}}) \\ &= h(h(x_3, x_1, \boldsymbol{\theta}_{\mathbf{1,2}}), h(x_2, x_1, \boldsymbol{\theta}_{\mathbf{1,1}}), \boldsymbol{\theta}_{\mathbf{2,1}}) \end{aligned}$$

$\boldsymbol{\theta}_{\mathbf{1,1}}, \boldsymbol{\theta}_{\mathbf{1,2}}, \boldsymbol{\theta}_{\mathbf{2,1}}$は，$C_{2,1}(\cdot)$, $C_{3,1}(\cdot)$, $C_{23|1}(\cdot)$のコピュラパラメータベクトル

C-vine copulaからの乱数生成

- 乱数生成手順

1) 互いに独立に，区間[0,1]の一様分布に従う乱数 w_1, w_2, w_3を生成する.

2) 1)で生成した一様乱数を用い

$$u_1 = w_1$$

$$u_2 = h^{-1}(w_2, w_1, \boldsymbol{\theta}_{\mathbf{1,1}})$$

$$u_3 = h^{-1}(h^{-1}(w_3, h(h^{-1}(w_2, w_1, \boldsymbol{\theta}_{\mathbf{1,1}}), w_1, \boldsymbol{\theta}_{\mathbf{1,1}}), \boldsymbol{\theta}_{\mathbf{2,1}}), w_1, \boldsymbol{\theta}_{\mathbf{1,2}})$$

想定したコピュラに対するh関数を用いて導出可能に

により，1組の乱数(u_1,u_2, u_3)を得る.

C-vine copulaからの擬似乱数生成

- SASプログラム例

煩雑さを避けるため，FCMPプロシジャを用いて
h関数とh逆関数を定義

```
proc fcmp outlib=work.func.my_funcs;
   function norm_h(・・・);
    ・・・;
   endsub;
   function invnorm_h(・・・);
    ・・・
   endsub;
   ・・・
run;
```

2変量正規コピュラのh関数を定義

2変量正規コピュラのh逆関数を定義

他コピュラのh関数とh逆関数を定義

C-vine copulaからの擬似乱数生成

- SASプログラム例

  3変数のC-vineに基づく擬似乱数を生成するためのSASマクロ%m3dcvinesimを用意

```
%macro m3dcvinesim(
   seed,          /*シード*/
   ndraws,        /*生成したい乱数の個数*/
   outuniform,/*生成した乱数のSASデータセット名*/
   family,        /*ペアコピュラのタイプ*/
   par,           /*コピュラパラメータ(除自由度)*/
   par2           /*自由度. tコピュラ使用時のみ*/
);
```

C-vine copulaからの擬似乱数生成

- SASプログラム例

3つの反復doステートメントにより1組の乱数を生成

```
do i=2 to 3;
 m_v(i,1)=m_w(i);
  do k=i-1 to 1 by -1;
    m_v(i,1)= invnorm_h(・・・);
  end;
  m_x(i)=m_v(i,1);
  do j=1 to i-1;
     m_v(i,j+1)=norm_h(・・・));
  end;
end;
```

iをインデックスとするループ内で2つのループを

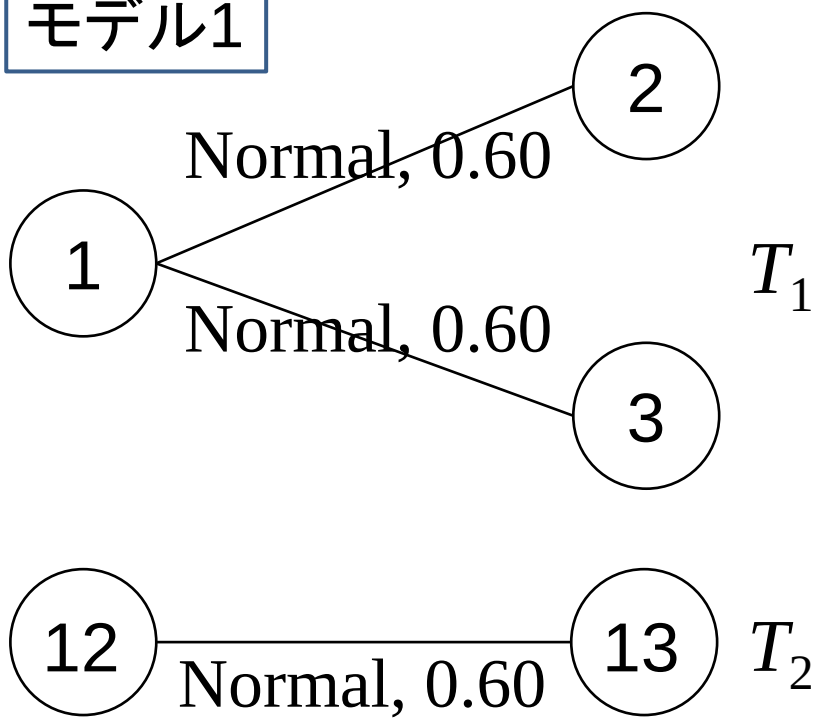
kをインデックスとするループ

jをインデックスとするループ

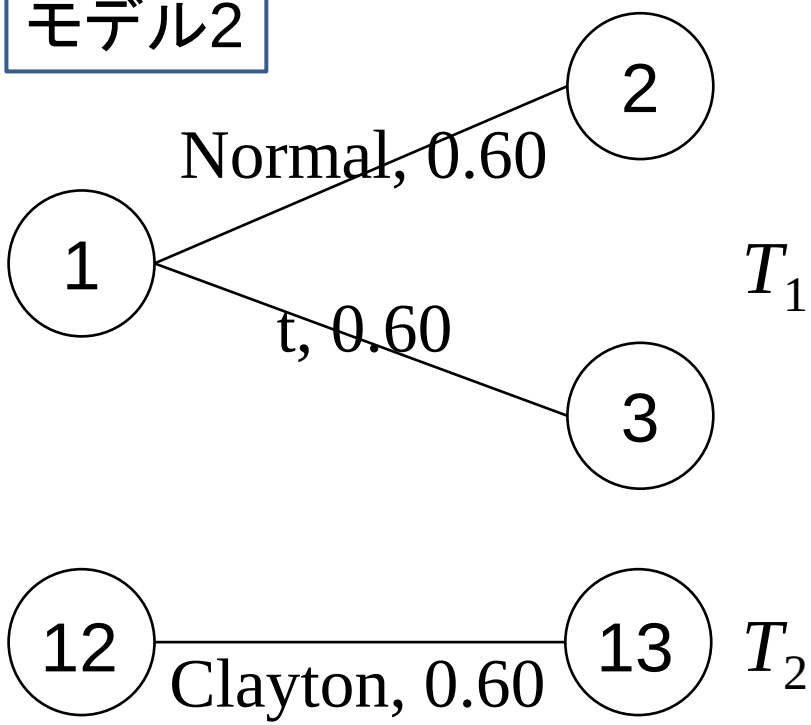
C-vine copulaからの擬似乱数生成

- 適用例

モデル1



モデル2

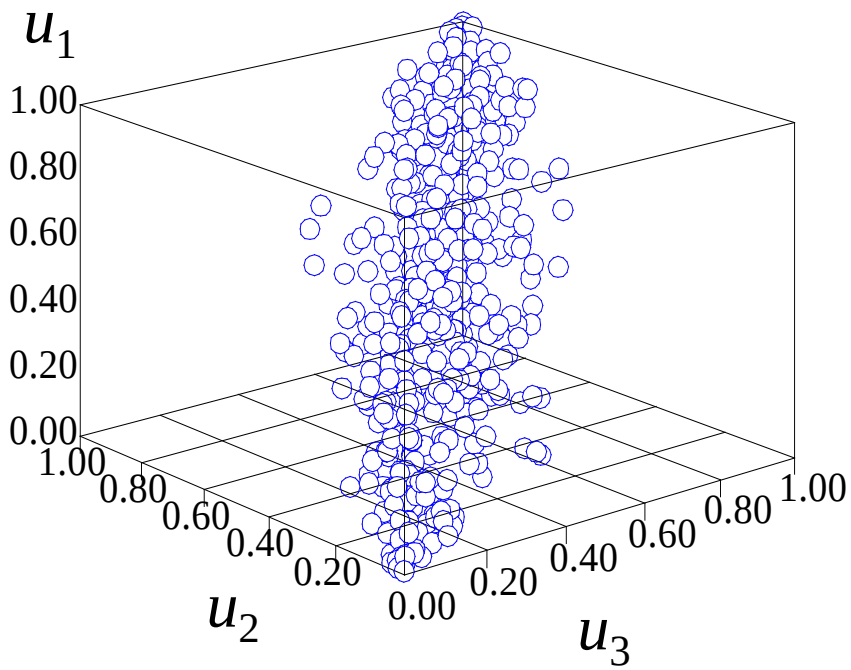


ケンドールのτが全て0.60になるよう，コピュラパラメータを設定

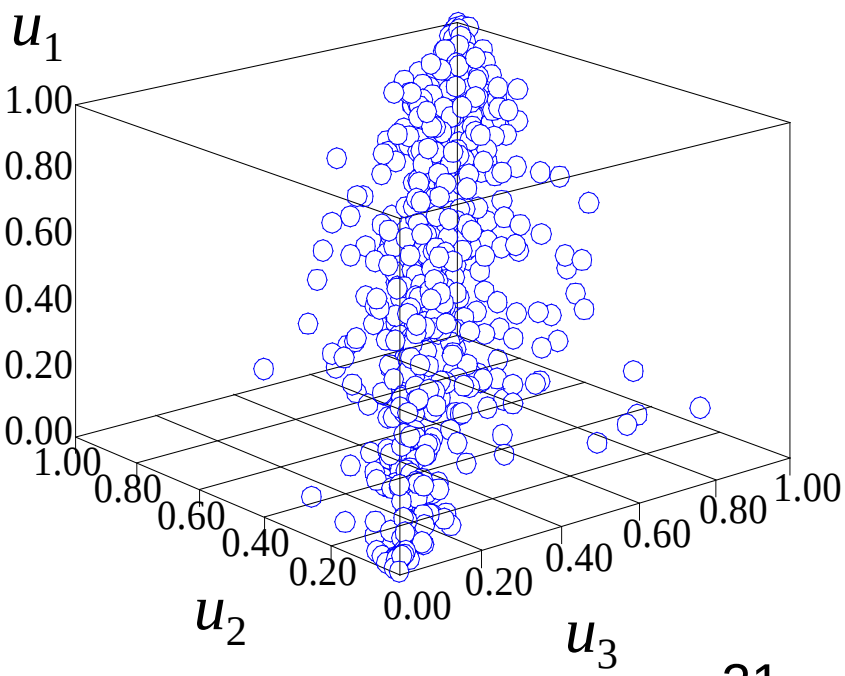
C-vine copulaからの擬似乱数生成

- 適用例
  - 生成した擬似乱数500個の3次元プロット

モデル1



モデル2



まとめ

Regular vineの特別なクラスであるC-vineについて，定義，乱数生成方法を簡潔に紹介

- 同時分布関数の構成

  ペアコピュラと周辺分布の積

- 乱数の生成

  h関数とその逆関数を用いて表記

- SASを用いた擬似乱数の生成

  FCMPプロシジャにより必要な関数を定義

参考文献

[1]Nelson, R.B. (2006). An Introduction to Copulas. New York: Springer.

[2]塚原英敦. (2008). 接合分布関数（コピュラ）の理論と応用. 日本統計学会HP 版. http://park.itc.u-tokyo.ac.jp/atstat/jss75shunen/Vol3.pdf（閲覧日：2015年6月21日）.

[3]矢田真城, 浜田知久馬. (2014). SASを用いたコピュラに従う擬似乱数の生成. SASユーザー総会論文集, 527–538.

[4]SAS Institute Inc. (2011). SAS/ETS 9.3 User's Guide. The COPULA Procedure. Cary, NC: SAS Institute Inc.

[5]SAS Institute Inc. (2013). SAS/ETS 13.1 User's Guide. The COPULA Procedure. Cary, NC: SAS Institute Inc.

[6]Bedford, T., Cooke, R.M. (2002). Vines- a new graphical model for dependent random variables. The Annals of Statistics, 30, 1031-1068.

[7]宮川雅巳. (1997). グラフィカルモデリング. 朝倉書店.

参考文献

[8]Kurowicka, D. and R. M. Cooke. (2004). Distribution - free continuous Bayesian belief nets. In Fourth International Conference on Mathematical Methods in Reliability Methodology and Practice, Santa Fe, New Mexico.

[9]Joe, H. (1996). Families of m-variate distributions with given margins and m(m-1)/2 bivariate dependence parameters. In L. R¨uschendorf, B. Schweizer and M.D.Taylor (eds.), Distributions with Fixed Marginals and Related Topics, 28, 120-141.

[10]Aas, K., C. Czado, A. Frigessi, and H. Bakken. (2009). Pair-copula constructions of multiple dependence. Insurance: Mathematics and Economics, 44, 182-198.

[11]Yan, J. (2007). Enjoy the Joy of Copulas: With a Package copula. Journal of Statistical Software, 21, 1-27.

[12]Brechmann, E. C., and U.Schepsmeier. (2013). Modeling dependence with C- and D-vine copulas: The R-package CDVine. Journal of Statistical Software, 52, 1-27.

[13]Lewandowski, D., Kurowicka, D. and Joe,H. (2009). Generating random correlation matrices based on vines and extended onion method. Journal of Multivariate Analysis, 100, 1989-2001.